

Data Quality Perspectives from the ESIP Information Quality Cluster

David Moroni₁(David.F.Moroni@jpl.nasa.gov),
Hampapuram “Rama” Ramapriyan₂, Ge Peng₃

1. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
2. NASA Goddard Space Flight Center and Science Systems and Applications, Inc.
3. NOAA’s Cooperative Institute for Climate and Satellites - North Carolina (CICS-NC) and NOAA’s National Centers for Environmental Information (NCEI)

Presented at the 2019 E2SIP C3DIS ‘19 Workshop, Canberra, Australia, 8 May 2019.

Acknowledgements: These activities were carried out across multiple United States government-funded institutions (noted above) under contracts with the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA). Government sponsorship acknowledged. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

- ESIP Information Quality Cluster
 - About ESIP IQC
 - Overview of IQC Objectives
 - Many players around the world
 - Data and Info Quality Perspectives
 - Maturity Models Supporting each Stage
 - Fostering Efforts within each Perspective
- Collaboration with the ESDSWG Data Quality WG
 - Operational Solutions Master List
 - Current and Forthcoming Publications of Standards and Practices

- Vision
 - Become **internationally recognized** as an **authoritative and responsive information resource** for guiding the implementation of **data quality standards and best practices** of the science data systems, datasets, and data/metadata dissemination services.
- Closely connected to Data Stewardship Committee
- Open membership (as with all Collaboration Areas in ESIP)

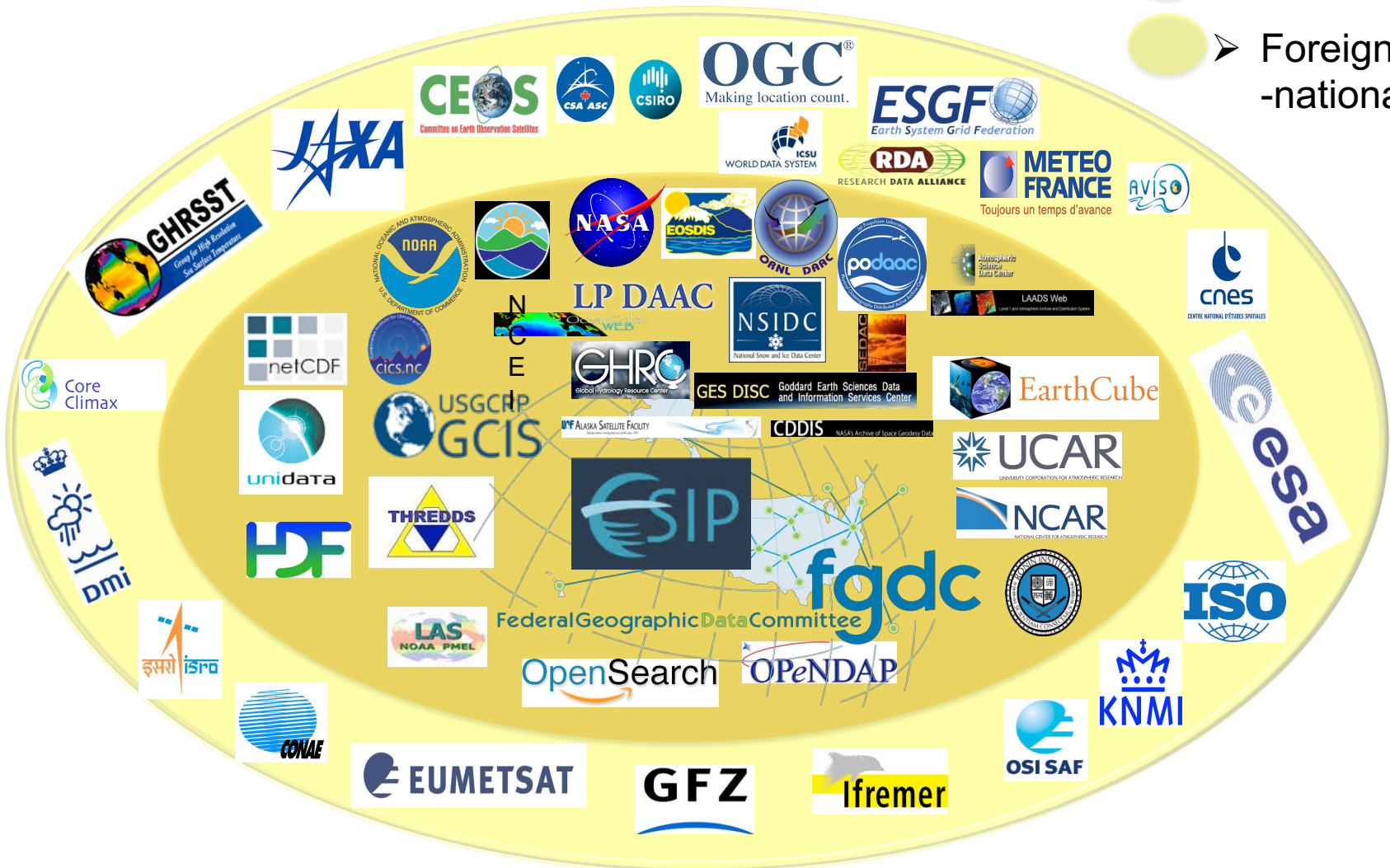
ESIP Information Quality:

http://wiki.esipfed.org/index.php/Information_Quality

- Share Experiences.
- Actively evaluate best practices and standards for data quality from the Earth science community.
- Improve collection, description, discovery, and usability of information about data quality in Earth science data products.
- Consistently provide guidance to data managers and stewards on the implementation of data quality best practices and standards as well as for enhancing and improving data maturity.
- Support:
 - Data producers with information about standards and best practices for conveying data quality; provide mentoring as needed
 - Data providers/distributors/intermediaries establish, improve, and evolve mechanisms to assist users in discovering, understanding, and applying data quality information properly.

Many Players Around the World

-  ➤ US
-  ➤ Foreign/Inter-national



- Scientific quality
 - Accuracy, precision, uncertainty, validity and suitability for use (fitness for purpose) in various applications
- Product quality
 - How well the scientific quality is assessed and documented
 - Completeness of metadata and documentation, provenance and context, etc.
- Stewardship quality
 - How well data are being managed, preserved, and cared for by an archive or repository
- Service Quality
 - How easy it is for users to find, get, understand, trust, and use data
 - Whether archive has people who understand the data available to help users.
- Ramapriyan, H K, Peng G, Moroni D, Shie C-L, **Ensuring and Improving Information Quality for Earth Science Data and Products**. D-Lib Magazine, 23 (7/8), July/August 2017, DOI: <https://doi.org/10.1045/july2017-ramapriyan>

Perspective: Based on different data product lifecycle stages

Define/Develop/Validate

Science

Produce/Assess/Deliver

Product

Maintain/Preserve/Disseminate

Stewardship

Use/User Service

Service

- Accuracy
- Precision
- Uncertainty
- fitness for purpose

- how well the product has been produced and assessed;
- Completeness of product metadata and documentation

- how well the data are being managed, preserved, and stewarded;
- How well are metadata and documentation for access

- how well the data are being serviced;
- user support
- customer engagement

Ramapriyan, H K, Peng G, Moroni D, Shie C-L, **Ensuring and Improving Information Quality for Earth Science Data and Products**. D-Lib Magazine, 23 (7/8), July/August 2017, DOI: <https://doi.org/10.1045/july2017-ramapriyan>

Maturity Models Supporting each Stage



Define/Develop/Validate
Science

Produce/Assess/Deliver
Product

Maintain/Preserve/Disseminate
Stewardship

Use/User Service
Service

Science
Maturity Matrix

Product
Maturity Matrix

Stewardship
Maturity Matrix

Service
Maturity Matrix

Zhao et al. (2016)
Thorne et al. (2015)

Bates and Privette
(2012)

Peng et al.
(2015)

NCEI MM-Serv WG
(under-development)

EUMETSAT (2013)

*Scientifically
sound and utilized*

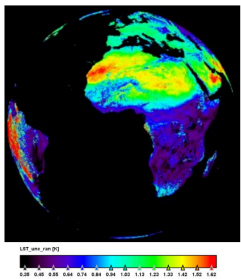
*Fully documented
and transparent*

*Well-preserved
and integrated*

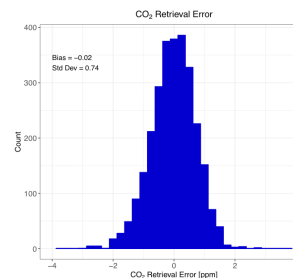
*Readily obtainable
and usable*

(Peng 2018, *Data Science Journal*)

- Science maturity models: GAIA-CLIM (Thorne et al. 2015), NOAA STAR Algorithm Maturity (Reed 2013; Zhou et al. 2016), CORE-CLIMAX Production System Maturity (EUMETSAT, 2013).
- Convened sessions on the topic of Earth science data uncertainty: ESIP 2017 Summer Meeting, AGU Fall 2017 Meeting, ESIP 2018 Summer Meeting AGU Fall 2018 Meeting.
- Internationally collaborated white paper entitled “Understanding and Communicating Uncertainty in Earth Science Data Informatics” (in draft). Est. pub. June 2019.
- Monthly IQC-organized telecons featuring experts on data calibration/validation, uncertainty quantification/characterization, quality assurance, and codification of quality information and metadata.



Land Surface
Temperature
Uncertainty (Merchant
et al. 2018)

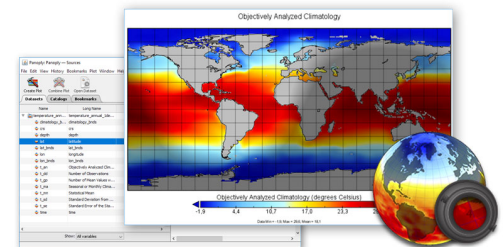


CO2 Retrieval Error
(Hobbs et al. 2018)

- Product maturity models: NOAA CDR Maturity Matrix (Bates and Privette 2012), CORE-CLIMAX Production System Maturity (EUMETSAT, 2013).
- Metadata standards and models promoting interoperability: Climate and Forecast (CF), ACDD, ISO 8601/19115-x/19157-x.
- Metadata standards and models for preservation and provenance: ISO 19165-2 (Rama et al. 2019).
- Data Model Cluster (Davis, 2019):
http://wiki.esipfed.org/index.php/Data_Model
- Monthly IQC-organized telecons featuring experts on data production, metadata formatting/validation, interoperability, and provenance.



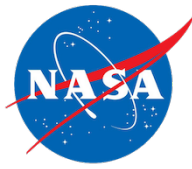
unidata



- Stewardship maturity models: NCEI/CICS-NC Stewardship Maturity (Peng et al. 2015), CEOS WGISS Data Management and Stewardship Maturity (WGISS DMSMM 2017), WMO Stewardship Maturity for Climate Data (SMM-CD Working Group, 2018).
- Metadata models for preservation, such as ISO-19165-2 (Rama et al. 2019).
- Leveraging open-source technologies and standards:
 - ESIP Semantic Technologies (SemTech) Committee (McGibbney et al., 2019), http://wiki.esipfed.org/index.php/Semantic_Technologies
 - ESIP Documentation Cluster (Gordon and Bugbee, 2019), http://wiki.esipfed.org/index.php/Category:Documentation_Cluster
 - ESIP Web Services Cluster (Gallagher, 2019), <http://wiki.esipfed.org/index.php/WebServices>
 - ESIP Usability Cluster (Hou and Langseth, 2019), <http://wiki.esipfed.org/index.php/Usability>
- Monthly organized IQC telecons featuring all of the above topics.

- Service maturity models: NSIDC Level of Services (Duerr et al. 2009), NCEI Tiered Scientific Data Stewardship Services (Peng et al., 2016a), GCOS ECV Data and Information Access Matrix, Global Ocean Observing System (GOOS) Framework, NCEI Data Monitoring and User Engagement Maturity (Arndt and Brewer, 2016).
- Keeping up with Big Data challenges: Velocity, Variety, Veracity.
- Leveraging input from other ESIP clusters:
 - Agriculture and Climate (Hoebelheinrich and Teng, 2019), http://wiki.esipfed.org/index.php/Agriculture_and_Climate
 - Disaster Lifecycle (Moe and Jones, 2019), <http://wiki.esipfed.org/index.php/Disasters>
 - Data to Decisions (Wee, 2019), http://wiki.esipfed.org/index.php/Data_to_Decisions
- Monthly organized IQC telecons featuring all of the above topics.

About ESDSWG DQWG

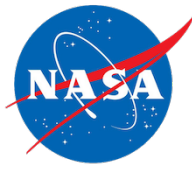


- One of NASA's Earth Science Data System Working Groups (ESDSWG).
- Formed at the annual meeting of the ESDSWG in 2014 as a result of interest expressed by the ESDIS Project and MEaSUREs investigators.
- Mission Statement
 - Evaluate and make recommendations to the ESDIS Project and HQ's Earth Science Data Systems (ESDS) Program for improvements in capturing, representing and enabling the use of data quality information describing accuracy, precision, uncertainty and applicability ("fitness for use") stewardship in the NASA Earth science domain.

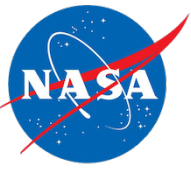
<https://earthdata.nasa.gov/community/>

ESDSWG: [earth-science-data-system-working-groups-esdswg](https://earthdata.nasa.gov/community/earth-science-data-system-working-groups-esdswg)

Operational Solutions Master List



- Intended to identify operational solutions (26) relevant to the Implementation strategies identified by the DQWG.
- <https://wiki.earthdata.nasa.gov/x/2pASBg>
- Solutions can either be software, documentation, or standards/practices.
- Solutions cover the following implementation categories:
 - Data Quality Information (representation/dissemination)
 - Facilitate Data Center and Provider/PI Communication
 - Metadata Creation
 - Standards Compliance Checking and Reporting
 - Guidance and Instruction
 - User Services
 - Knowledgebase



- **ESDS-RFC-031**: *Data Management Plan Template for Distributed Active Archive Centers (Convention)*
- **ESDS-RFC-032**: *Data Management Plan Template for Data Processing Systems (Convention)*
- **ESDS-RFC-033**: *Comprehensive Data Quality Recommendations for Data Producers and Distributors (Suggested Practice)*
- **ESDS-RFC-034**: *High-Priority Data Quality Recommendations for Data Producers and Distributors (Suggested Practice)*
- **Submitted to ESO (under review)**: *Assessment of Recommended Data Quality Software Products (Technical Note)*
- **Preparing Submission to ESO**: *Report of Data Call Pilot Study (Technical Note)*



<https://earthdata.nasa.gov/about/esdis-project/esdis-standards-office-eso>

■ Primary Contacts:

- David Moroni (JPL, David.F.Moroni@jpl.nasa.gov)
- Yaxing Wei (ORNL DAAC, weiy@ornl.gov)
- H. K. “Rama” Ramapriyan (SSAI/GSFC – ESDIS, hampapuram.ramapriya@ssaihq.com)
- Ge Peng (NOAA/CICIS-NC/NCEI, ge.peng@noaa.gov)

■ Learn more about the IQC:

http://wiki.esipfed.org/index.php/Information_Quality

■ Access to DQWG Master List of Solutions:

<https://wiki.earthdata.nasa.gov/x/2pASBg>

■ Access to NASA’s ESO Standards and Technical Notes:

<https://earthdata.nasa.gov/about/esdis-project/esdis-standards-office-eso>